Why Do Pooled Forecasts Do Better Than Individual Forecasts Ex Post?

Diego Nocetti
Clarkson University

William T. Smith
The University of Memphis

Abstract

Pooled forecasts frequently outperform individual forecasts of economic time series. This paper shows that the introduction of model uncertainty into the formation of expectations can account for the regularity. We conjecture that agents learn in a Bayesian way, using an optimally designed combination of forecasts to form expectations. When these expectations alter the ex-post realization of the data generating mechanism the pooled forecast may dominate the best individual device.

We wish to thank two anonymous referees for helpful comments.
Citation: Nocetti, Diego and William T. Smith, (2006) "Why Do Pooled Forecasts Do Better Than Individual Forecasts Ex Post?." *Economics Bulletin*, Vol. 4, No. 36 pp. 1-7
Submitted: August 16, 2006. **Accepted:** October 25, 2006.
URL: http://economicsbulletin.vanderbilt.edu/2006/volume4/EB-06D80016A.pdf

1. Introduction

It is well known that pooled forecasts frequently outperform individual forecasts of economic time series *ex-post*, in the sense that the pooled forecast of alternative models provides a smaller mean-squared forecast error (MSFE) than *any* of the individual models' MSFEs[1]. Bernanke and Boivin (2003), for example, show that pooled forecasts of inflation and unemployment do as well or better than the Federal Reserve's Greenbook at all horizons. Although several explanations have been proposed (see, e.g., Hendry and Clements (2002)), the reason for this regularity is still an open question. We provide a new explanation for this phenomenon.

Our point of departure is to relax the standard assumption that economic agents know the true data generating process of the economic variable. We conjecture that agents face model uncertainty and learn in a Bayesian way, using a combination of forecasts to form expectations. In particular, they deal with model uncertainty by practicing Bayesian-Model-Averaging (BMA) across a set of alternative models. Developed by Leamer (1978), BMA has found a range of applications in the economics literature[2]. In these applications it is the econometrician who faces model uncertainty, while the agents in the economy know the true model[3]. Our innovation is to assume that the agents themselves face model uncertainty, and resolve the problem as a Bayesian econometrician would.

Min and Zellner (1993) show that *ex-ante* forecasts pooled in a BMA fashion provide a lower expected squared error loss that any individual model when the forecaster faces model uncertainty. Intuitively, for an agent that is uncertain about the true model of the economy, selecting a forecast combination provides insurance against large forecast errors. However, this does not explain why pooled forecasts should do so well *ex-post*, *i.e.*, once the state of the economy is uncovered.

With our novel way of modeling expectations it is simple to see why pooled forecasts perform so well (*ex-post*): If agents form their expectations with BMA, their forecasts feed back endogenously to affect the equilibrium, altering the *ex-post* realization of the data generating mechanism.

Our method is closely related to that of Evans and Honkapohja (2001). They assume that agents consider one, possibly misspecified, model and form forecasts using least squares regressions. Since the forecast functions affect the state of the economy, their approach, like ours, is *self-referential*. However, our approach has two distinctive characteristics. First, we incorporate into the agents' forecasting problem uncertainty about

[1] Bates and Granger (1969) pioneered this literature. Hendry and Clements (2003) provide an excellent theoretical examination on the current state of the literature on forecasts combination and its advantages over the use of single forecasts. See also Clemen (1989) for an annotated bibliography.

[2] Areas of application include monetary policy (Brock, Durlauf and West (2003, 2004), Cogley and Sargent (2005)), economic growth (Doppelhofer, Miller, and Sala-i-Martin (2000) and Fernandez, Ley, and Steel (2001)), Finance (Avramov (2002)), and forecasting of exchange rates (Wright (2003a)) and inflation (Wright (2003b)).

[3] Brock, Durlauf and West (2003, 2004) and Cogley and Sargent (2005) analyze model uncertainty using Bayesian methods from the perspective of a policy-maker.

the structure of the economy, a feature that is present in most economic environments. For instance, our working example includes the stylized fact that the variables that enter a monetary policy rule are uncertain. Second, although adaptive expectations might be reasonable, the forecast functions of misspecified models are inherently arbitrary. Instead, the advantage of using BMA is that it is the *ex ante* optimal way of forecasting and, therefore, it renders consistent the existence of multiple models.

2. Expectations formation with model uncertainty

A simple AD-AS model (Lucas, 1973; Evans and Honkapohja, 2001 p.29) provides a convenient platform to depict expectations formation under model uncertainty[4]. Suppose that the economy consists of the aggregate supply and aggregate demand curves

$$\begin{aligned} y_t &= \bar{y} + \gamma(p_t - E_{t-1}p_t) + v_t \\ m_t &= p_t + y_t \end{aligned} . \tag{1}$$

The money supply rule is

$$m_t = \bar{m} + \varepsilon_t + \eta X_{t-1} \tag{2}$$

v_t and ε_t are unobserved white noise shocks, $E_{t-1}p_t$ denotes the expectation of p_t conditional on information available at $t-1$, and X_{t-1} is an exogenous *observable* shock. The equilibrium price satisfies

$$p_t = \mu + \lambda E_{t-1}p_t + \delta X_{t-1} + \xi_t \tag{3}$$

where $\mu = (\bar{m} - \bar{y})/(1+\gamma)$, $\lambda = \gamma/(1+\gamma)$, $\delta = \eta/(1+\gamma)$, and $\xi_t = (\varepsilon_t - v_t)/(1+\gamma)$.

For notational convenience let $\bar{m} = \bar{y}$. The rational expectations equilibrium (REE) would be

$$p_t = \frac{\delta X_{t-1}}{1-\lambda} + \frac{\xi_t}{1-\lambda} \equiv p^e_{t,RE} + \frac{\xi_t}{1-\lambda}. \tag{4}$$

This will serve as a useful benchmark.[5]

---

[4] It should be noted, however, that our analysis encompasses a large class of models. For instance, most of the applications of econometric learning presented by Evans and Honkapohja (2001) can be readily extended to incorporate model uncertainty.

[5] A semantic note: We are using the term "rational expectations" in the narrow, traditional sense where agents use all available information and *know the correct model*. However, the BMA forecast is, by the Min and Zellner Theorem (1993) the "rational" way of forming expectations *when the model is not known* (see Proposition 1 below). We only highlight the REE because (1) it will help develop the intuition for our results, and (2) in some circumstances (Section 5) the BMA forecast converges asymptotically to the rational expectations forecast.

Agents consider a set of K alternative models of the economy that differ in the exogenous shock postulated to drive the money supply rule. The reduced form of model *i* $(i = 1,.....,K)$ is

$$p_t = \alpha_i X_{t-1,i} + \xi_{t,i}. \quad (5)$$

We assume that one of the models, model *k*, includes the actual exogenous shock in (2).

In the face of this uncertainty agents form expectations using Bayesian-Model-Averaging. First, agents assign a prior probability $p(M_i), i = 1,...,K$ to each model. Then, they assign a prior distribution to the unknown parameters, given the model probability distributions $p(\alpha_i | M_i)$. Finally, the data D is generated given the previous two distributions $p(D | \alpha_i, M_i)$. Conditioning in the observed data yields for each model the posterior probabilities, which can be interpreted as the probability that the ith model is the true model:

$$\pi_i = p(M_i | D) = \frac{p(D|M_i)p(M_i)}{\sum_{i=1}^{K} p(D|M_i)p(M_i)} \quad (6)$$

where $p(D|M_i) = \int p(D|\alpha_i, M_i)p(\alpha_i|M_i)d\alpha_i$ is the marginal likelihood of the ith model.

The forecast of any model i is then

$$p_{t,i}^e = \hat{\alpha}_i X_{t-1,i}. \quad (7)$$

The BMA forecast is

$$p_{t,BMA}^e = \sum_{i=1}^{K} \pi_i p_{t,i}^e, \quad (8)$$

The point forecast is thus the forecast for each individual model weighted by its posterior probability distribution.

The use of BMA is not arbitrary, but is a natural procedure for optimizing agents facing model uncertainty. Specifically, Min and Zellner (1993) demonstrate that BMA provides the minimum expected square error[6]. Therefore

Proposition 1. *If the forecaster seeks to minimize the expected square forecast error, where the expectations are taken over the model space, it is optimal to form expectations using BMA.*

[6] Madigan and Rafterty (1994) also show that BMA is better than any single forecast when forecast ability is measured by a logarithmic scoring rule.

Although we omit the proof, which is in Min and Zellner (1993), the intuition of the result is that, for an agent that is uncertain about the true model of the economy, selecting a forecast combination provides insurance, i.e. she can never be "too wrong". The proposition refers to *ex ante* forecasts. It is a normative assertion: agents *should* use a pooled forecast if they face model uncertainty.

3. Equilibrium

We will show that BMA can provide the best forecasts not only *ex ante*, but also *ex post*. Why? Unlike Min and Zellner (1993), the forecasters here do not operate outside the system. The forecasts now feed back endogenously to affect the equilibrium price: agents form their expectations based on (8), which in turn affects the actual price generated in accordance with the true model. As in the literature on adaptive expectations (see e.g. Evans and Honkapohja (2001)), the use of BMA by forecasters makes the data generating process endogenous, i.e. a self-referential system.

Consider the realized equilibrium price. Substituting the BMA forecast (8) into the reduced form equation (3), we see that the price is

$$p_t = \lambda \sum_{i=1}^{K} \pi_i p_{t,i}^e + \delta X_{t-1} + \xi_t. \tag{9}$$

It follows that the forecast error for the BMA model is

$$p_t - p_{t,BMA}^e = (\lambda - 1) \sum_{i=1}^{K} \pi_i p_{t,i}^e + \delta X_{t-1} + \xi_t. \tag{10}$$

To develop some intuition for what this means, it is useful to re-write (10) using the equilibrium price of the Rational Expectations model [equation (4)] as a benchmark:

$$p_t - p_{t,BMA}^e = (1 - \lambda) B_{RE,BMA} + \xi_t, \tag{11}$$

where $B_{RE,BMA} \equiv p_{t,RE}^e - p_{t,BMA}^e$ is the bias of the BMA forecast relative to the REE forecast. Equation (11) says that the BMA forecast error is proportional to the average bias of all the models relative to the REE forecast. In other words, the BMA forecast error depends upon the bias in the equilibrium price (relative to REE) *caused by the fact that agents use BMA forecasting.*

Similarly, consider the forecast error of any model j (including the REE model),

$$\begin{aligned} p_t - p_{t,j}^e = \xi_{i,j} &= \lambda \sum_{i=1}^{K} \pi_i p_{t,i}^e + \delta X_{t-1} + \xi_t - p_{t,j}^e \\ &= B_{RE,j} - \lambda B_{RE,BMA} + \xi_t. \end{aligned} \tag{12}$$

where $B_{RE,j} \equiv p_{t,RE}^e - p_{t,j}^e$ is the bias of model j relative to the REE forecast. Suppose that model j predicts a price higher than in the REE forecast, so that $B_{RE,j} < 0$. Then Equation

(12) tells us that the mean forecast error of model j will be lower if the BMA forecast also forecasts a price in excess of the RE forecast. In other words, the "intrinsic" error of model j can be offset if the use of BMA forecasting raises the actual equilibrium price, relative to the REE equilibrium.

4. When does BMA do best *ex post?*

The mean-square forecast error of BMA is[7]

$$E\left(\left(p_t - p^e_{t,BMA}\right)^2 \middle| \pi_i, X_{t-1,i} \ \forall i\right) \equiv \phi_{BMA} = (1-\lambda)^2 B^2_{RE,BMA} + \sigma^2_\xi \tag{13}$$

The mean-square forecast error of any model j is

$$E\left(\xi^2_{t,j} \middle| X_{t-1,j}\right) \equiv \phi_j = B^2_{RE,j} - 2\lambda B_{RE,j} B_{RE,BMA} + \lambda^2 B^2_{RE,BMA} + \sigma^2_\xi . \tag{14}$$

The BMA forecast does better than any single model if $\phi_{BMA} < \phi_j$. When will this happen? In the appendix we show

Proposition 2. *If* $1/2 > \lambda$ *then BMA does better than model j ex post when*

$$\left|B_{RE,j}\right| < \left|B_{RE,BMA}\right| < \left|\frac{B_{RE,j}}{1-2\lambda}\right|. \tag{15}$$

If $1/2 < \lambda$ *then BMA does better than model j if*

$$\left|B_{RE,BMA}\right| > \left|\frac{B_{RE,j}}{1-2\lambda}\right|. \tag{16}$$

To provide some intuition for this result we provide a graphical analysis of the case where $B_{RE,j} < 0$ and $\lambda < 1/2$, that is, when model j over-predicts the price relative to the REE and the weight attached to the BMA forecast in the equilibrium price is low.

Figure 1 shows the mean-square forecast errors of BMA and model j as functions of the bias of the BMA forecast (relative to REE). First consider the mean square error of BMA, ϕ_{BMA}. It is an increasing function of the absolute value bias of the BMA forecast: the larger the bias of BMA relative to REE, the less accurate is the BMA forecast.

Now consider the mean square forecast error of model j, ϕ_j. Recall that the forecast error of model j depends upon both its "intrinsic" bias *and* the bias in the equilibrium induced by the use of BMA. On the one hand, BMA forecasting reduces the accuracy of model j by creating a bias in the equilibrium price, relative to REE [this is the $B^2_{RE,BMA}$ term in (14)]. On the other hand, if BMA causes the equilibrium price to be high (relative to REE), then this may offset the "intrinsic" bias of model j [this is the interaction

---

[7] Note that, since the exogenous shocks $X_{t-1,i}$ are observable at the time when the forecast is made, the only stochastic term in the one-step-ahead forecast error is ξ_t .

term $2\lambda B_{RE,j} B_{RE,BMA}$ in (14)]. For sufficiently low BMA bias, the latter effect will dominate the former, and BMA will do better than model j.

5. Example and Asymptotic behavior

We now present a simple case that provides more transparent results in terms of the classic linear regression model and allows us to study convergence. Specifically, we assume that the observable exogenous shocks are IID and uncorrelated and that the representative agent holds diffuse prior beliefs on a_i and diffuse priors on the model space.

Without trying to contend generality, the assumption of diffuse priors is an interesting case since the Bayesian parameter estimates reduce to the classical ordinary least squares estimates and the BMA weights can be approximated by the fit of the least squares regressions. Specifically, the BMA weights are BIC adjusted likelihoods: $\pi_i \approx \frac{e^{(-0.5*BIC_i)}}{\sum_i^K e^{(-0.5*BIC_i)}}$ (see e.g. Rafterty, 1995). We believe that this process, which Doppelhofer, Miller, and Sala-i-Martin (2000) call *Bayesian averaging of classical estimates*, gives more realism to our hypothesis that agents use BMA to form expectations.

Similar to the previous analysis, the DGP proceeds as follows:

1. Given an initial price, agents use least squares to form the forecast rules $p_{t,i}^e = \hat{a}_i X_{i,t-1}$.
2. Agents calculate the BMA weights for model *i* given the *BIC* adjusted likelihoods.
3. The equilibrium price is generated.
4. The observed actual price leads to a change in expectations and a change in the value of the parameters and the weights attached to each model, which in turn affects the price, and so on.

Let $\hat{a}_{i,t}$ be the least squares estimate using the prices observed up to period *t* and the exogenous variable *i* up to period t-1, i.e. $\hat{a}_{i,t} = \left(\sum_{l=t-1}^{t-1-N} X_{i,l}^2 \right)^{-1} \left(\sum_{l=t-1}^{t-1-N} X_{i,l} p_{l+1} \right) \equiv \hat{\sigma}_t^{i,p} \big/ \hat{\sigma}_{i,t}^2$. The actual price is generated in accordance with (9), which now satisfies

$$p_{t+1} = \lambda \sum_{i=1}^{K} \frac{e^{(-0.5*BIC_{i,t})}}{\sum_i^K e^{(-0.5*BIC_{i,t})}} \left(\hat{\sigma}_t^{i,p} \big/ \hat{\sigma}_{i,t}^2 \right) X_{i,t} + \delta X_t + \xi_t . \qquad (17)$$

This equation permits establishing the restrictions over the parameters of the models under which BMA does better than any model *j* (equations 15 and 16).

In addition, the following proposition establishes the asymptotic behavior of (17)

Proposition 3. *Under the assumptions above,* $\hat{\sigma}_t^{i,p} \Big/ \hat{\sigma}^2_{i,t} \rightarrow 0$ *and* $\dfrac{e^{(-0.5*BIC_{i,t})}}{\sum_i^K e^{(-0.5*BIC_{i,t})}} \rightarrow 0$ $\forall i \neq k$, *i.e. except for the model that includes the exogenous shock in the actual policy rule. For model k we have* $\hat{\sigma}_t^{k,p} \Big/ \hat{\sigma}^2_{k,t} \rightarrow \dfrac{\delta}{1-\lambda}$ *and* $\dfrac{e^{(-0.5*BIC_{k,t})}}{\sum_i^K e^{(-0.5*BIC_{k,t})}} \rightarrow 1$, *that is, the REE.*

Proof. see appendix.

In other words, the BMA forecast converges asymptotically to the rational expectations forecast. The intuition of this result is that the model that includes the exogenous shock in the actual policy rule will, on average, do better than the other models. This, in turn, implies that agents will tend to increase the BMA weights attached to this model over time. As they do so, the least squares estimates of the model that uses the exogenous shock in the monetary policy rule converge to the REE.

6. Conclusion

Proposition 2 actually applies to *any* pooled forecast: A pooled forecast could dominate individual forecasts simply because forecasters chose to use it. This would explain why non-BMA pooled forecasts, such as numerical averages, also do better than individual forecasts. The advantage of focusing on BMA is that it is the *ex ante* optimal way of forecasting in the presence of model uncertainty. Further, BMA forecasting allows us to investigate the convergence properties of the self-referential system.

Appendix

Proof of Proposition 2

First consider the mean-square forecast error of BMA as a function of $B_{BMA,k}$:

$$\phi_{BMA}\left(B_{RE,BMA}\right) = (1-\lambda)^2 B_{RE,BMA}^2 + \sigma_\xi^2 \qquad \text{(A.1)}$$

Plotting ϕ_{BMA} as a function of $B_{RE,BMA}$ yields a parabola that reaches a minimum at σ_ξ^2.

Next consider the mean-square forecast error of model j:

$$\phi_j\left(B_{RE,BMA}\right) = \left(B_{RE,j} - \lambda B_{RE,BMA}\right)^2 + \sigma_\xi^2. \qquad \text{(A.2)}$$

Note that

$$\phi_j(0) = B_{RE,j}^2 + \sigma_\xi^2 > 0. \qquad \text{(A.3)}$$

$$\phi_j^{/} = -2\lambda\left(B_{RE,j} - \lambda B_{BMA,RE}\right) \qquad \text{(A.4)}$$

$$\phi_j^{//} = 2\lambda^2 > 0 \qquad \text{(A.5)}$$

(A.2), (A.4) and (A.5) imply that ϕ_j reaches a minimum of σ_ξ^2 at $\hat{B}_{RE,BMA} = \dfrac{B_{RE,j}}{\lambda}$.

Finally, the mean-square forecast errors of the two models are equal when

$$0 = (1-2\lambda)B_{RE,BMA}^2 + 2\lambda B_{RE,BMA}B_{RE,j} - B_{RE,j}^2. \qquad \text{(A.6)}$$

The roots to this equation are $B_{j,RE}$ and $-B_{j,RE}/(1-2\lambda)$. This identifies four different constellations of parameter values, depending upon the signs of $B_{RE,j}$ and $1-2\lambda$, from which proposition 2 follows.

Proof of Proposition 3.

Our argument follows Bray and Savin's (1986) analysis of least squares convergence to the REE.

Fix $\hat{a}_{i,t-1}$, and π_i. By the strong law of large numbers $n\left(\sum_{l=t-1}^{t-1-n} X_{i,l}^2\right)^{-1} \to \left[Var\left(X_i\right)\right]^{-1}$ almost surely and $\dfrac{1}{n}\left(\sum_{l=t-1}^{t-1-n} X_{i,l}\, p_{l+1}\right) \to \lambda\pi_i\hat{a}_{i,t-1}Var\left(X_i\right)$. This implies $\hat{a}_{i,t} = \lambda\pi_i\hat{a}_{i,t-1} \quad \forall i$ except for model *k*. If we now allow $\hat{a}_i$ to evolve over time, and since $\lambda\pi_i < 1$, this estimates converge to zero. Furthermore, since the coefficients attached to any model besides the RE model converge to zero, BIC_i, and as a result the BMA weights, also converge to zero

Using a similar argument, for model k we have $\hat{a}_{k,t} = \lambda\pi\hat{a}_{k,t-1} + \delta X_{t-1}$. In the limit this converges to $\dfrac{\delta X_{t-1}}{1-\lambda\pi}$; but, since for this model $\pi = 1$ we obtain the REE.

References

Avramov, D., (2002) “Stock Return Predictability and Model Uncertainty” *Journal of Finance* 64, 423-458.

Bates, J., and C. Granger (1969) “The combination of forecasts” *Operations Research Quarterly* 20, 451-68.

Bernanke, B., and J. Boivin (2003) “Monetary Policy in a Data-Rich Environment” *Journal of Monetary Economics* 50, 525-546.

Bray, M., and N. Savin (1986) “Rational expectations equilibria, learning, and model specification” *Econometrica* 54, 1129-60.

Brock, W., Durlauf, S., and K. West (2003) “Policy Analysis in Uncertain Economic Environments (with discussion)” *Brookings Papers on Economic Activity*, 1, 235-322, 2003.

Brock, W., Durlauf, S., and K. West (2004) “Model Uncertainty and Policy Evaluation: Some Theory and Empirics.” NBER Working Paper 10916.

Clemen, R., (1989) “Combining forecasts: A review and annotated bibliography” *International Journal of Forecasting* 5, 559–583.

Cogley, T., and T. Sargent (2005) “The Conquest of U.S. Inflation: Learning and Robustness to Model Uncertainty” *Review of Economic Dynamics* 8, 528-563 .

Doppelhofer, G., Miller, R., and X. Sala-i-Martin (2000) “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach.” NBER Working Paper 7750.

Evans, G.W., and S. Honkapohja (2001) *Learning and Expectations in Macroeconomics*. Princeton, NJ: Princeton University Press.

Fernandez. C., Ley E., and M. Steel (2001) “Model Uncertainty in Cross-Country Growth Regressions” *Journal of Applied Econometrics* 16, 563-76.

Hendry, D. F., and M. Clements (2002) “Pooling of forecasts” *Econometrics Journal* 5, 1-26.

Hendry, D. F., and M. Clements (2003) “Economic Forecasting: Some Lessons From Recent Research” *Economic Modelling* 20, 301-329.

Leamer, E., (1978) *Specification Searches*. New York: Wiley

Lucas, R.E. (1973). “Some international evidence on output-inflation tradeoffs” *The American Economic Review* 63, 326-334

Madigan, D., and A.E. Rafterty (1994) “Model Selection and Accounting for model uncertainty in graphical models using Occam’s window” *Journal of the American Statistical Association* 89, 1535-1546

Min, C., and A. Zellner (1993) “Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates” *Journal of Econometrics*, 56, 89-118

Rafterty, A.E. (1995) “Bayesian model selection in social research (with discussion)” In: Peter V. Marsden (ed.), *Sociological Methodology*. Cambridge, MA: Blackwell.

Wright, J., (2003a) “Bayesian Model Averaging and Exchange Rate Forecasting.” Federal Reserve Board International Finance Discussion Papers 779.

Wright, J., (2003b) “Forecasting US Inflation by Bayesian Model Averaging.” Federal Reserve Board International Finance Discussion Papers 780.

Figure 1

$B_{RE,j} < 0 \qquad 1-2\lambda < 0$